Title	Symbolic-Numeric Optimization for Estimation of Parameters in a Biological Kinetic Model(Computer Algebra : Design of Algorithms, Implementations and Applications)
Author(s)	ORII, SHIGEO; HORIMOTO, KATSUHISA; ANAI, HIROKAZU
Citation	数理解析研究所講究録 (2006), 1514: 66-73
Issue Date	2006-09
URL	http://hdl.handle.net/2433/58682
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録
1514巻 2006年 66-73

Symbolic-Numeric Optimization for Estimation of Parameters in a Biological Kinetic Model

折居 茂夫
SHIGEO ORII
富士通株式会社
FUJITSU LTD [*]

堀本 勝久
KATSUHISA HORIMOTO
産業技術総合研究所・生命情報科学研究センター
COMPUTATIONAL BIOLOGY RESEARCH CENTER, AIST [†]

穴井 宏和
HIROKAZU ANAI
(株) 富士通研究所/(独) 科学技術振興機構
FUJITSU LABORATORIES LTD / CREST, JST [‡]

Abstract

We have been studying a symbolic-numeric optimization for estimation of parameters in biological kinetic models by quantifier elimination (QE), in combination with numerical simulation methods. The optimization method was applied to a model for the inhibition kinetics of HIV proteinase with ten parameters and nine variables. We apply this optimization procedure to three sets of observed data and obtain kinetic parameters by using only one point of each set of the data.

1 Introduction

Many methods for local and global optimization have been developed to model and simulate the global network of biological molecules in a cell [1, 2], and some simulators based on various optimization methods have also been designed (e.g. [3]). In the optimization methods, the estimation of kinetic parameters plays a key role in the development of kinetic models, which, in turn, promotes functional understanding at the system level, for example, in several biological pathways [4, 5]. An answer to the estimation of kinetic parameters is our symbolic-numeric optimization which combines symbolic QE with numerical simulation [6, 7]. In this paper, firstly, we show our procedure of the optimization for the inhibition kinetics of HIV proteinase [8], which includes an enhanced procedure of the offset computation. Secondly, we show that the kinetic parameters for three sets of observed data can be estimated by using only one point of each set of the data.

---

[*] orii@strad.ssg.fujitsu.com
[†] horimoto@cbrc.jp
[‡] anai@jp.fujitsu.com

2 MATERIALS AND METHODS

2.1 Mathematical Framework

Problem: In this paper, we consider the following fitting problem: the biological kinetic model analyzed here is of the form:

$$\dot{x}_i = v_i(X, K) \tag{1}$$

where $X = \{x_1, \cdots, x_{n_i}\}$ is a set of variables, and $K = \{k_1, \cdots, k_{n_j}\}$ is a set of parameters. The problem is to fit the parameters K of the model to the observed data $\tilde{X} = \{\tilde{x}_i^t\}$ for , $i = 1, \cdots, n_x$, $t = 0, 1, \cdots, n_{\tilde{x}_t}$ under the following additional conditions:

(i) Conservation laws: $h_i(X) = 0$

(ii)Variable ranges: $x_i \in D_i$, where $D_i = [a, b]$, $a, b \in \mathbb{R} \cup \{\infty\}$.

Basic Formula Here we set up the leading formula of this paper. As mentioned above, we have the following constraints Ψ with error variables ei from kinetic models: $\Psi \equiv \wedge_i \psi_i$, where $\psi_i = \dot{x}_i - v_i(X, K) + e_i = 0$. For the error variables we introduce a new variable, e_max, which means the magnitude of the error variables: $|e_i| \leq e_{max}$. Moreover, for the variables whose observed data is given, we consider the following objective conditions: $X_l^{(t)} - \tilde{X}_l^{(t)} = 0$, to achieve fitting. Then the " basic formula " is given as

$$F(\dot{X}, X, K, e_{max}, e_i) \equiv (\Psi \wedge h_i(X) = 0 \wedge x_i \in D_i \wedge |e_i| \leq e_{max} \wedge X_l^{(t)} - \tilde{X}_l^{(t)} = 0). \tag{2}$$

We apply our symbolic-numeric approach to formulas derived by slightly modifying the basic formula according to various purposes.

2.2 Optimization Procedure

We explain the concrete procedure of symbolic-numeric optimization, which consists of six parts as illustrated in Figure 1.

(1) **Numerical simulation** First we prepare simulation data for $\dot{x}_i$ and x_i, for which we lack observed data, by performing a numerical simulation of the kinetic models.

1. Set initial conditions $\tilde{X}^{(0)}$ and starting values for unknown parameters $\tilde{K}^{(0)}$ as follows: $\tilde{X}^{(0)} \equiv \{\tilde{x}^{(0)} | i = 1, \cdots, n_x\}$ and $\tilde{K}^{(0)} = K_1^{(0)} \cup K_2^{(0)}$, where $\tilde{K}_1^{(0)} \equiv \{k_1^{(0)}, \cdots, k_j^{(0)}\}$ are starting values, and $\tilde{K}_2^{(0)} \equiv \{k_{j+1}^{(0)}, \cdots, k_{n_j}^{(0)}\}$ are given fixed parameters.
2. By numerical simulation of the kinetic model (1), we obtain a time series for x_i and $\dot{x}_i$: $X_i^{(t)} = \{x_i^{(t)} | i = 1, \cdots, n_i, t = 0, 1, \cdots, n_t\}$ and $\tilde{\dot{X}}_i^{(t)} = \{\tilde{\dot{x}}_i^{(t)} | i = 1, \cdots, n_i, t = 0, 1, \cdots, n_t\}$.

(2) **Formulation** After choosing some variables from X, we call them " focusing variables ", Y, and substitute observed/simulated data into the remaining variables:

1. Choose a subset Y of $X : Y \subseteq X$.
2. Substitute $\dot{X}$, $X \setminus Y$, in F by the values of $\tilde{\dot{X}}, \tilde{X}$ at a time point t: $\dot{X}_i \leftarrow \tilde{\dot{X}}_i^{(t)}, \dot{X}_i \leftarrow \dot{X}_i^{(t)}$, where $x_i \in \tilde{\dot{X}}, X \setminus Y$. Then we denote the new formula as $F'(Y, K_1, e_{max}, e_i)$. We note by performing a QE computation for the formula, $\exists Y \exists K_1 \exists e_{max} \exists e_i (F')$,

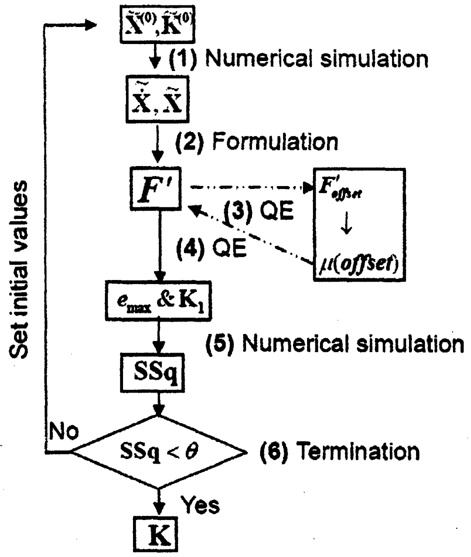

Figure 1: Flowchart of symbolic-numeric optimization. The variables and values are enclosed by the boxes, and the procedures are numbered corresponding to the description in the text.

(3) Computation of offset by QE Observed data often contain an offset. Therefore, we must first determine the offset value. Here we consider the case that the offset appears linearly. For the sake of simplicity, we assume that only $\tilde{x}_1$ has an offset. Let F'_{offset} be the formula obtained by putting $\tilde{x}'_1 - offset$ into $\tilde{x}_1^{(t)}$ of F', where $offset$ is a variable for offset. By performing QE for $\exists X \exists K_1 \exists e_{max} \exists e_i (F'_{offset})$, we obtain the quantifier-free formula $\pi(offset)$, which stands for the feasible ranges of $offset$. Then we substitute the minimum value of the offset for the variable offset in F', and we denote it again by $F'(Y, K_1, e_{max}, e_i)$.

(4) Estimation of *emax* by QE First, we use QE to find the magnitude of *emax* as small as possible. By computing QE for $F'(Y, K_1, e_i)$, we obtain a quantifier-free formula $\pi(emax)$ describing the feasible ranges of *emax*. Next, we put the minimum value of e_{max} into e_{max} in F', and denote the resulting formula as $F''(Y, K_1, e_i)$.

And Estimation of K_1 by QE We obtain a quantifier-free formula $\tau(K_1)$ describing the feasible ranges of K_1 by computing QE for $Y \exists e_i (F'')$, Actually, the feasible ranges of K_1 are usually sufficiently narrow intervals (e.g., about 10^{-6}) to choose an appropriate specific value of K_1.

(5) Computation of sum of squares (SSq) by numerical simulation We estimate the goodness-of-fit for the obtained parameter valuesK_1 from the feasible ranges ofK_1 in terms of SSq.

1. Set initial conditions $\tilde{X}^{(0)}$ and K_1.
2. Perform numerical simulation of kinetic model (1).
3. Compute $SSq : SSq = \sum_t (x_1^{(t)} - \tilde{x}_1^{(t)})^2$.

(6)**Termination** If SSq is smaller than a specific level θ, output K. Otherwise, set new initial values and go to (1).

2.3 Biological Model

We analyzed a model for the inhibition kinetics of HIV proteinase [8], as shown in Figure 2. The proteinase monomer (M) is inactive, but the enzyme (E) is active in the dimeric form. The dimer catalyzes the conversion of the substrate (S) to the product (P). The inhibitor (I) is competitive for the substrate and the product, and the inhibitor-binding enzyme is irreversibly deactivated (EJ). In the model, there are ten parameters and nine variables. According to the previous studies [8, 9], five parameters ($k_{11}, k_{12}, k_{21}, k_{41}, k_{51}$) are given, and the remaining five unknown parameters ($k_{22}, k_3, k_{42}, k_{52}, k_6$), two initial values ($E_{init}, S_{init}$) and the offset of the fluorimeter are estimated by the present method. The experimental data of the product $[P]$, which are composed of 300 data points measured from 0 to 3600 seconds, were downloaded from a web site (http://www.gepasi.org/tutorials/opt/hivfit.html).

$$\begin{array}{llll}
M+M & \rightleftharpoons & E & k_{11}(\longrightarrow), k_{12}(\longleftarrow) \\
S+E & \rightleftharpoons & ES & k_{21}, k_{22} \\
ES & \rightarrow & E+P & k_3 \\
E+P & \rightleftharpoons & EP & k_{41}, k_{42} \\
E+I & \rightleftharpoons & EI & k_{51}, k_{52} \\
EI & \rightarrow & EJ & k_6
\end{array}$$

Figure 2: Kinetic model for the inhibitor of HIV proteinase. The start values for ten parameters and the initial values for nine variables [9] are as follows: $k_{11} = 0.1, k12 = 10^{-4}, k_{21} = 100, k_{22} = 300, k_3 = 10, k_{41} = 100, k_{42} = 500, k_{51} = 100$, $k_{52} = 0.1, and k_6 = 0.1; \tilde{x}_1 = 0, \tilde{x}_2 = 0.004, \tilde{x}_3 = 25.0, \tilde{x}_4 = 0, \tilde{x}_5 = 0, \tilde{x}_6 = 0, \tilde{x}_7 = 0.003, \tilde{x}_8 = 0, and \tilde{x}_9 = 0$.

3 RESULTS

First, we will describe the practical procedure for parameter optimization in the kinetic model for HIV proteinase, and then we will evaluate the optimized parameters by using only one point of the observed data.

3.1 Procedure for Optimizing Parameters in HIV inhibition Model

To perform the numerical simulation (in (1) of 2.2),K_1 and K_2, are defined as the five unknown parameters and the five given parameters, and the nine variables are allocated to $[P]$, $[E]$, $[S]$, $[ES]$, $[M]$, $[EP]$, $[I]$, $[EI]$, and $[EJ]$. Then we set the start value $\tilde{K}^{(0)}$ and the initial value $\tilde{X}^{(0)}$. The start values for ten parameters and the initial values for nine variables are cited from the previous study [9] (see the legend in Figure 2). Also, the two initial values, Einit and Sinit, are changed within a limited range with reference to the previous studies [8, 9]: 31 discrete values for ($[E]$ =0.00350, 0.00355, $\cdots$, 0.00500) and 13 values for ($[S]$ =23.0, 23.5, $\cdots$, 29.0). The focusing variables Y (in (2) of 2.2) are simply obtained by the symbolic computation with QE from the relationship between X andK_1 in the model. In the inequality $v_i(X, K)\Delta t + x_i^t \geq 0$, the elimination of Δt by QE outputs five inequalities including five parameters: $100*[E]*[I]-k52*[EI]-k6*[EI] > 0, 100*[E]*[I]-k52*[EI] > 0, 100*[E]*[P]-k42*[EP]-k3*[ES] < 0, 100*[E]*[P]-k42*[EP] > 0, and 100*[E]*[S]-k22*[ES]-k3*[ES] > 0$. Among the five unknown

Table 1: Goodness of fit with optimized parameters by symbolic-numeric method. (a) is in the case of I=0, which means no inhibition. (b) is in the case of I=0.0015. (c) is in the case of I=0.003. *Itr* is the iterations number of the symbolic-numeric optimization.

(a)

time	*Itr*	E_{init}	S_{init}	k_{22}	k_3	k_{42}	k_{52}	k_6	*SSq*
336	1	0.00350	25.5	215.2	9.972	1082	-	-	0.00758
984	1	0.00350	25.5	140.3	9.907	573.8	-	-	0.00824
1848	1	0.00350	25.5	144.2	9.935	544.2	-	-	0.00951
Mendes & Kell	-	0.004389	24.79	201.1	7.352	1171	1.31E+04	3.00E+4	0.00347
Kuzmic	-	-	-	179.7	9.46	1117	0.0831	0.1224	-

(b)

time	*Itr*	E_{init}	S_{init}	k_{22}	k_3	k_{42}	k_{52}	k_6	*SSq*
336	1	0.00360	23.0	195.5	9.909	1878	0.103	0.0972	0.0321
984	1	0.00350	23.5	111.9	9.971	870.1	0.105	0.0960	0.0320
1848	1	-	-	-	-	-	-	-	-
Mendes & Kell	-	0.004537	26.79	201.1	7.352	1171	1.31E+04	3.00E+4	0.00985
Kuzmic	-	-	-	179.7	9.46	1117	0.0831	0.1224	-

(c)

time	*Itr*	E_{init}	S_{init}	k_{22}	k_3	k_{42}	k_{52}	k_6	*SSq*
336	1	0.00495	27.5	250.9	9.776	1296	0.103	0.0969	0.00895
984	1	0.00470	28.0	162.8	9.980	1134	0.102	0.0982	0.00795
1848	1	0.00390	29.0	38.67	9.990	1342	0.101	0.0986	0.04250
	2	0.00465	28.0	149.5	9.805	1241	0.110	0.0970	0.00835
Mendes & Kell	-	0.005470	26.79	201.1	7.352	1171	1.31E+04	3.00E+4	0.00513
Kuzmic	-	-	-	179.7	9.46	1117	0.0831	0.1224	-

For reference, the values related to the present optimization are also cited from previous studies [8, 9].

parameters in the above five inequalities, $[P]$ is included in the objective function, and $[S]$ is a large value relative to the other variables in the reaction molecules. Except for the last three inequalities including $[P]$ and $[S]$, only $[EI]$ appears in the terms related to the unknown parameters in the first two inequalities. Thus, the focusing variables Y are defined as $[P]$, $[S]$, and $[EI]$ in the present model. All symbolic computations by QE in this study are performed by REDUCE (ver. 3.7) (http://www.uni-koeln.de/REDUCE/). In addition, the conservation laws in the present model are obtained by Gepasi [3], a tool for estimating the kinetic flux in a given model, as follows: $h_1(X) = [S]+[ES]+[P]+[EP]-S_{init} = 0$ and $h_2(X) = [M] + 2[E] - 2[S] - 2[P] + 2[EI] + 2[EJ] - (2E_{init} - 2S_{init}) = 0$.

The computation of offset by QE ((3) of 2.2) is realized by eliminating all of the variables by QE, except for $offset$ in F'_{offset}. By the elimination, the following three equations composed of the initial values and the observed values are obtained in the present model: $\mu(offset) = [E] + [EI] + [EJ] + [M]/2 - [Einit] - [Sinit] = 0, [EP] + [ES] + [P] + [S] - [Sinit] = 0, and offset + 3/125[P] - x_1 = 0$.

From the last two equations, we can obtain $offset = \tilde{x}_1 - 3/125 * ([Sinit] - [S] - [EP] - [ES])$.

By considering the properties of the kinetic model, this equation can be approximated with the observed data. In the initial state, $[EP]$ and $[ES]$ are much less than $[S]$, and as the reaction proceeds, $[S]$ decreases steadily. Therefore, $[Sinit] >> [S] - [EP] - [ES]$ at a steady state. Thus, we can obtain $offset = (\tilde{x}_1)_{steady} - 3/125 * [Sinit]$, where $(\tilde{x}_1)_{steady}$ is a value of $\tilde{x}_1$. In the present study, we used the value of $(\tilde{x}_1)_{steady}$ at $t = 3600$ as the value of $(\tilde{x}_1)_{steady}$.

Using F' of Y and the offset obtained above, we can estimate emax andK_1 by QE (in (4) of 2.2). Note that 403 sets of emax andK_1 are obtained by the corresponding sets of E_{init} and S_{init}. Since the fitting of simulated data strongly depends on the initial values, we further simulate numerically E_{init} and S_{init} within the above ranges of E_{init} and S_{init}; by a standard technique of the bisection method, E_{init} and S_{init} for each set of emax andK_1 are estimated to minimize the SSq that is calculated for 300 values of $[P]$ (in (5) of 2.2). Finally, we obtain a set of e_{max},K_1, E_{init} and S_{init} by selecting a minimum SSq among the 403 SSq's.

To judge whether the loop in Figure 1 terminates or not (in (6) of 2.2), the minimum of SSq's is compared with the threshold θ. The threshold is set to 0.04 in the present study.

3.2 Observed Data Fitting with the Optimized Parameters

The optimized parameters with the six sets of observed data are listed in Table 1, together with the iteration number, the goodness of fit measured by SSq, the initial values of E_{init} and S_{init}, and the offset. In addition, the fittings of simulated values to the observed data in six cases are described in Figure 3.

One of the remarkable features of the present fitting is that only one point of the observed data are sufficient to fit 300 data points with an SSq value of less than 0.03. The data point for the optimization is randomly chosen from 300 points of data, and all fittings attain the threshold by one or two iterations of the loop. In one of the six cases, two rounds of iterations were required, but the first fitting in the case agreed well with the observed data. This is partly because QE powerfully restricts the possible ranges of the parameters and the variables, and partly because the present model is simpler than that expected from the complex kinetics of ten parameters and nine variables. These points will be discussed in the following section.

Another feature is that the values of the parameters agree well with those in the previous studies [8, 9]. In particular, the highlighted parameters in this model, the inhibitor binding constant (k_{52}) and the deactivation rate constant (k_6), are about 0.10 and 0.097 in the six cases, which are similar values to the constants in one previous study [8]. In contrast, the constants are enormously large in the other previous study [9]. In comparison with both cases, the value in the latter case is unreasonably large for the analysis to be acceptable. Thus, the large dissociation and deactivation rate constants suggest that the potency of the inhibitor is overestimated in terms of the inhibitor reaction.

4 DISCUSSION

Two problems in the present optimization remain: the first of them is the choice of the observed data for the optimization, and the second of them is the choice of $(\tilde{x}_1)_{steady}$ in the offset computation. As for the first problem, the data showing a flat slope in the kinetic curve seem intuitively inadequate for

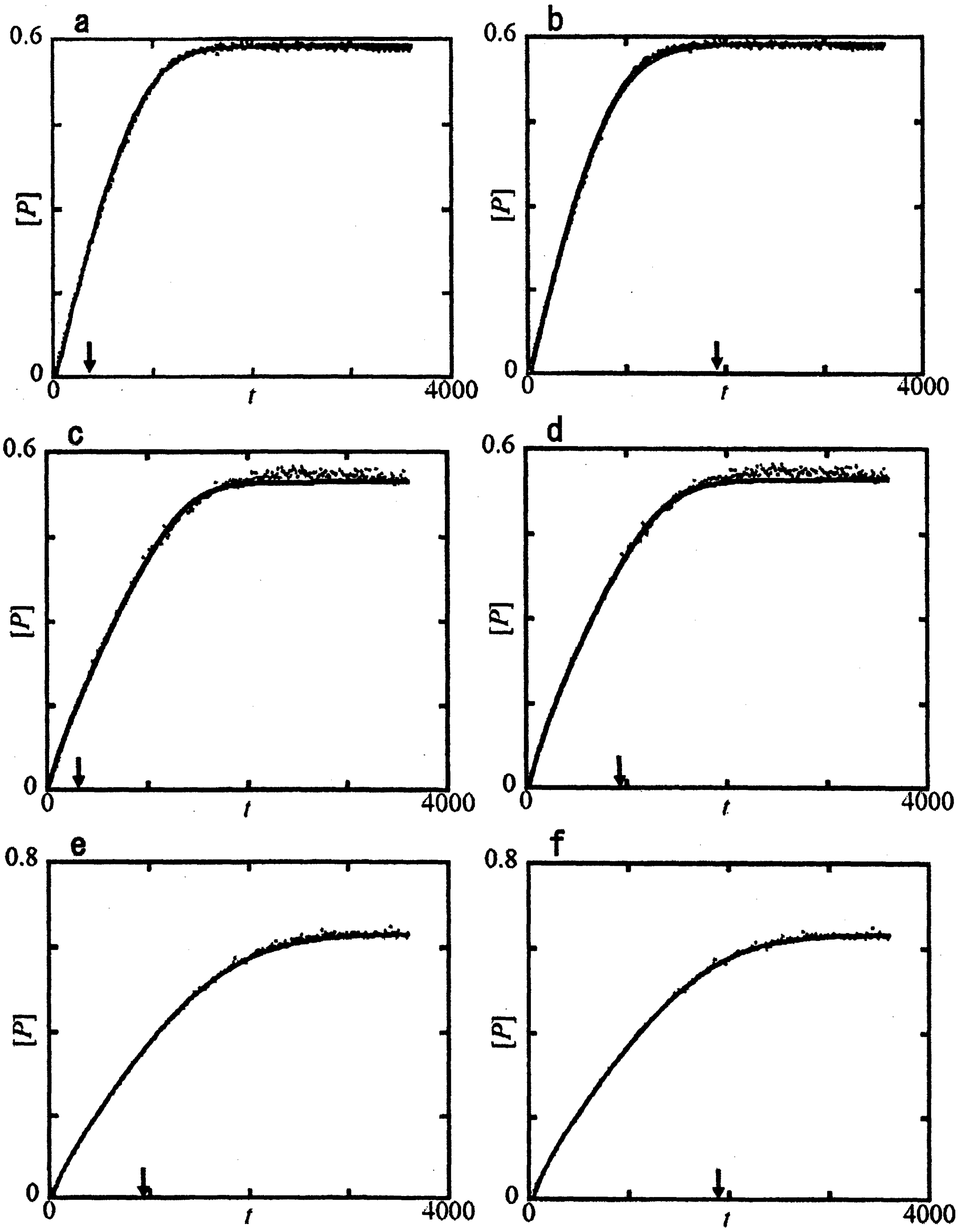

Figure 3: Fitting to observed data with optimized parameters. The amount of product $[P]$ is multiplied by a coefficient (0.024), according to [9]. The experimental data are denoted by the dots. **a** and **b** are in the case of $I = 0$. **a** is the carve of minimum SSq (=0.00758) and **b** is the carve of maximum SSq (=0.00951) in the (a) of Table 1. **c** and **d** are in the case of $I = 0.0015$. **c** is $SSq = 0.0321$ and **d** is $SSq = 0.0320$ in the (b) of Table 1. **e** and **f** are in the case of $I = 0.003$. **e** is the carve of minimum SSq (=0.00795) and **f** is the carve of maximum SSq (=0.00835) in the (c) of Table 1. The arrow of each figure denotes the time of the observed value used for the symbolic-numeric optimization: **a**, t =336; **b**, t =1848; **c**, t =336; **d**, t =984; **e**, t =984; **f**, t =1848.

the optimization. Indeed, by using the data of more than t =2500 in Figure 3, QE frequently outputs ' false ' ; this means no parameter and variable spaces for the initial conditions in F'. Any data, except for those in the steady states, may possibly output ' true ' for the optimization by QE. As for the second problem, fluctuation of data in steady states is the cause of large SSq value. In the case of $I = 0$ and $I = 0.003$ (see **a**, **b**, **e**, **f** of Figure 3), there is small fluctuation in the steady states. However, large fluctuation appears in the case of $I = 0.0015$ (see **c**, **d** of Figure 3) and $(\tilde{x}_1)_{steady}$ of $t = 3600$ is a lower value of the steady state. The estimated curve is adjusted to the point is the problem. A rule of data selection is required to attain more good SSq value.

Acknowledgements

We would like to express our gratitude to Dr. Taku Takeshima for his kind assistance. One of the authors (K. H.) was partly supported by a Grant-in-Aid for Scientific Research on Priority Areas "System Genomics" (grant 18016008) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] P. Mendes, and D.B. Kell, "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation", Bioinformatics, Vol. 14, 1998, pp. 869-883.

[2] C.G. Moles, P. Mendes, and J.R. Banga, "Parameter estimation in biochemical pathways: a comparison of global optimization methods", Genome Res., Vol. 13, 2003, pp. 2467-2474.

[3] P. Mendes, and D.B. Kell, "MEG (Model Extender for Gepasi): a program for the modelling of complex, heterogeneous, cellular systems", Bioinformatics, Vol. 17, 2001, pp. 288-289.

[4] M.H. Hoefnagel, M.J.C. Starrenburg, D.E. Martens, J. Hugenholtz, M. Kleerebezem, I.I. Van Swam, R. Bongers, H.V. Westerhoff, and J. Snoep, "Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modeling, metabolic control and experimental analysis", Microbiology, Vol. 148, 2002, pp. 1003-1013.

[5] I. Swameye, T.G. Muller, J. Timmer, O. Sandra, and U. Lingmuller, "Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling", Proc. Natl. Acad. Sci. USA., Vol. 100, 2003, pp. 1028-1033.

[6] S. Orii, H. Anai, and K. Horimoto, "Symbolic-numeric estimation of parameters in biochemical models by quantifier elimination", Proceedings of the 2005 International Joint Conference of InCoB, AASBi and KSBI, 2005, pp. 272-277.

[7] 折居茂夫，穴井宏和，堀本勝久，"Symbolic-Numeric Optimization for Kinetic Models, - An application to bioinformatics field -", 京都大学数理解析研究所研究集会 CA-ALIAS 2004 講究録 No.1456, pp40-48, 2005.

[8] P. Kuzmic, "Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase", Anal. Biochem., Vol. 237, 1996, pp. 260-273.

[9] P. Mendes, and D.B. Kell, "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation", Bioinformatics, Vol. 14, 1998, pp. 869-883.